



CrossMark

An analysis of four methodologies for estimating highway capacity from ITS data

Zhao Li · Rilett Laurence

Received: 12 December 2014 / Revised: 14 April 2015 / Accepted: 16 April 2015 / Published online: 21 May 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract With the recent advent of Intelligent Transportation Systems (ITS), and their associated data collection and archiving capabilities, there is now a rich data source for transportation professionals to develop capacity values for their own jurisdictions. Unfortunately, there is no consensus on the best approach for estimating capacity from ITS data. The motivation of this paper is to compare and contrast four of the most popular capacity estimation techniques in terms of (1) data requirements, (2) modeling effort required, (3) estimated parameter values, (4) theoretical background, and (5) statistical differences across time and over geographically dispersed locations. Specifically, the first method is the maximum observed value, the second is a standard fundamental diagram curve fitting approach using the popular Van Aerde model, the third method uses the breakdown identification approach, and the fourth method is the survival probability based on product limit method. These four approaches were tested on two test beds: one is located in San Diego, California, U.S., and has data from 112 work days; the other is located in Shanghai, China, and consists of 81 work days. It was found that, irrespective of the estimation methodology and the definition of capacity, the estimated capacity can vary considerably over time. The second finding was that, as expected, the different approaches yielded different capacity results. These estimated capacities varied by as much as 26 % at the San Diego test site and by 34 % at the Shanghai test site. It was also found that each of the methodologies has

advantages and disadvantages, and the best method will be the function of the available data, the application, and the goals of the modeler. Consequently, it is critical for users of automatic capacity estimation techniques, which utilize ITS data, to understand the underlying assumptions of each of the different approaches.

Keywords Capacity estimation method · Van Aerde model · Breakdown identification · PLM

1 Introduction

The Highway Capacity Manual (HCM) has been updated regularly (1965, 1985, 2000, and 2010) since it was first published in 1950 and its underlying theory has remained consistent [1–5]. While the HCM provides a uniform methodology for estimating the capacity for any highway in the U.S., many jurisdictions, both within the U.S. and outside the U.S., would prefer to use capacity values reflective of their own local conditions. With the recent advent of Intelligent Transportation Systems (ITS), and their associated data collection and archiving capabilities, there is now a rich data source for transportation professionals to develop capacity values for their own jurisdictions. Unfortunately, there is no consensus on the best approach for estimating capacity from ITS data.

While this study is concerned with estimating capacity, the problem is a subset of a much broader issue—how to identify the fundamental speed–flow–density relationship for a given facility. If the form of the underlying speed–flow–density fundamental diagram for a given facility is known, the capacity may be readily obtained given the appropriate empirical data. Needless to say the assumptions underlying the speed–flow–density function will affect the resulting capacity estimate.

Z. Li (✉)
University of Nebraska Lincoln, 2200 Vine St, 330P Whittier
Research Center, Lincoln 68503, USA
e-mail: li@huskers.unl.edu

R. Laurence
University of Nebraska Lincoln, 2200 Vine St, 262D Whittier
Research Center, Lincoln 68503, USA

This paper first estimates capacity concept using the (1) maximum method, (2) Van Aerde model, (3) breakdown identification, and (4) product limit method from ITS data collected in San Diego, California, U.S. and Shanghai, China. The values are then compared across time to examine the variability of capacity estimates as a function of location and methodology. In addition, the paper compares these capacity values with those obtained from the HCM. The paper concludes with a description of the advantages and disadvantages of each approach.

2 Literature review

2.1 Related work

The HCM 2010 [4] defines the capacity of a facility as “the maximum sustainable hourly flow rate at which persons or vehicles reasonably can be expected to traverse a point or a uniform section of a lane or roadway during a given time period under prevailing roadway, environmental, traffic, and control conditions”. The HCM capacity is a function of the free-flow speed. For example, if free-flow speed equals 110 km/h, the capacity is 2,400 pcphpl; if free-flow speed equals 80 km/h, the capacity is 2,200 pcphpl. This manual has been adopted in many jurisdictions around the world, mainly because it provides a single, deterministic value that represents average conditions. The HCM is based on speed–flow–density fundamental diagrams that were developed using empirical flow rate data collected across the U.S. for similar facilities. The flow rate at the apex of the speed–flow curve is regarded as the maximum throughput of the facility and thus is treated as the capacity [6].

While the use of the HCM is straightforward, identifying capacity from empirical observations is not. Consider the data shown in Fig. 1 which is from a detector site on the westbound I405 in San Diego California. Note that this site will be described in detail later in this paper. Intuitively, a

number of curve fitting algorithms could be used for estimating the speed–flow relationship in Fig. 1. The purple curve was developed using a standard generic curve fitting model, and it may be seen that the capacity for this example is 2,110 veh/h/ln. However, it may be seen that there are observed flows that are greater than this estimated capacity. A number of authors [7–9] have shown that even under “constant” conditions, both the maximum traffic flow and the capacity can vary over time and space. Other authors [10] have argued that the actual maximum observed flow rate of the roadway is the best measure of capacity because it: (1) is closer to the definition of capacity listed in the HCM and (2) does not assume a prior speed–flow–density relationship. Referring to the example in Fig. 1, the highest observed flow rate is 2,200 veh/h/ln, and this would represent the capacity according to the maximum capacity definition. The focus on this paper is to examine four popular methods of estimating capacity for ITS data similar to that shown in Fig. 1.

Minderhoud et al. [10] compared several capacity estimation methods including the headway method, bimodal distribution method, selected maxima method, the fundamental diagram method, and the on-line procedure estimation method. The comparison was done based on theoretical characteristics and no field data was used. The authors pointed out that there has been no comprehensive study of the validity and accuracy of these methods and, at the time the paper was written, this was an open research question. Similar work was conducted by Geistefeldt and Brilon [11] who compared stochastic methods for estimating capacity using empirical data. They found that the estimates varied over time, and that this should be accounted for in the estimation approach. In recent years a number of authors have developed new methods for estimating highway capacity using stochastic approaches [12, 13]. In general, these new approaches are based on existing popular methods such as model fitting, stochastic distribution, and breakdown-related methods that are examined in this paper. The accuracy of these new methods has not been compared empirically.

This paper compares both deterministic and stochastic methods using data from the same test beds. Specifically, four of the most popular methods are first introduced: Maximum method and Van Aerde method (deterministic) and breakdown method and PLM method (stochastic). Capacity values over time were estimated for each test site. The distribution of estimated capacity values were then compared statistically. Lastly, the four capacity estimation techniques were compared in terms of (1) data requirements, (2) modeling effort required, (3) estimated parameter values, (4) theoretical background, and (5) statistical differences across time and over geographically dispersed locations.

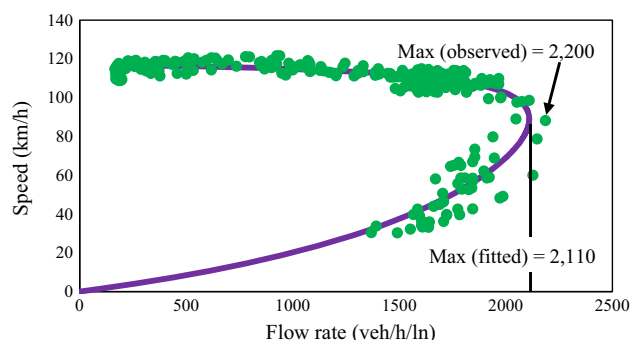


Fig. 1 Flow–speed diagram of westbound I405, San Diego, April 3rd, 2013 (5-min aggregation period)

2.2 Capacity estimation approaches for a single day

This paper will examine four of the most popular capacity estimation techniques for ITS data. A number of authors feel that capacity can change as a function of time and space, all else being equal, and therefore a particular focus of this paper will be on the techniques that examine the stochastic nature of capacity. Four of the most widely used methods (e.g., maximum method, fundamental diagram curve fitting using the Van Aerde model, breakdown identification method, and product limit method) were chosen for capacity estimation. A brief overview of the calibration and estimation procedures of the first three capacity methods are provided in this section.

2.2.1 Maximum capacity methodology

Arguably the easiest way to estimate capacity at a given location is to obtain the maximum flow rate from observed data measured over a given time period (e.g., a day), as shown in Eq. (1).

$$C_i = \max f_{i,\delta} \quad \forall \delta = 1, 2, \dots, N, \quad (1)$$

where, C_i denotes the maximum flow rate (e.g., capacity) over a given time period δ for location i ; δ denotes the time interval (e.g., 5 min); $f_{i,\delta}$ denotes the observed flow rate during time interval δ at location i ; and N denotes the number of time periods considered, e.g., $N = 288$ for $\delta = 5$ min.

When applying this method, the time interval δ will, by definition, affect the resulting capacity value C_i . All else being equal as the time interval δ increases, the capacity value C_i will decrease. For this reason, it is critical to define the time interval when presenting capacity values and to never compare capacity values that were developed from different time durations. Typical time durations range from 1–60 min. The HCM uses a 15-min time interval when defining capacity.

2.2.2 Van Aerde capacity methodology

The first and most famous deterministic speed–flow–density model was developed by Greenshields [14] and is based on the assumption of a linear speed–density relationship. Van Aerde [15] proposed a four-parameter model that provides more degrees of freedom to capture the range of behavior across different regimes and facility types. Note that this approach may still be considered a single regime traffic flow model. The Van Aerde model, which requires four input parameters, was designed specifically to be calibrated using empirical ITS data. While the original applications used field inductance loop data, any detector data, such as that from radar and video detectors, can be

used. The functional form of Van Aerde model is shown in Eq. (2).

$$C_i = \frac{u_i}{c_1 + \frac{c_2}{u_{f,i} - u_i} + c_3 u_i}, \quad (2)$$

where, C_i denotes the estimated capacity for location i ; u_i denotes the space mean speed (km/h) for location i ; $u_{f,i}$ denotes the free flow speed (km/h) for location i ; and, c_1 , c_2 , and c_3 denote the headway constant coefficients.

The model parameters are calculated using Eqs. (3–6).

$$m = \frac{2u_c - u_f}{(u_f - u_c)^2}, \quad (3)$$

$$c_1 = mc_2, \quad (4)$$

$$c_2 = \frac{1}{k_j \left(m + \frac{1}{u_f} \right)^2}, \quad (5)$$

$$c_3 = \frac{-c_1 + \frac{u_c}{q_c} - \frac{c_2}{u_f - u_c}}{u_c}, \quad (6)$$

where, u_c denotes the speed at capacity (km/h); q_c denotes the flow at capacity (veh/h); and k_j denotes the jam density (veh/km).

Once the model is calibrated, the capacity is identified as the maximum flow defined by the calibrated speed–flow curve. Because of its simplicity, this approach is relatively easy to program.

2.2.3 Breakdown capacity methodology

While widely used around the world, the general approaches described above (e.g., HCM, Van Aerde, etc.) have been criticized because they do not consider the stochastic nature of congestion and thus may be unsatisfactory for traffic operations applications [16, 17]. It has been argued that the point of traffic breakdown might be more appropriate for estimating capacity. Elefteriadou et al. [7, 18] showed that the traffic breakdown does not necessarily occur at the same volume level over different days and therefore that capacity should not be treated as a deterministic value. It should be noted that while the newly updated HCM 2010 acknowledges that capacity is stochastic, it does not identify a methodology to estimate the stochastic component of capacity.

In general, a breakdown is defined when the speed decrease between two consecutive time intervals exceeds a pre-specified threshold, and this lower speed is sustained for a predefined length of time [18]. A review of the literature indicates that the breakdown speed threshold and the congested time duration are location dependent. For example, Lorenz and Elefteriadou [19] defined breakdown as occurring when the average speed of all lanes drops below 90 km/h for a period of at least 5 min. Brilon et al. [17] used 70 km/h for their studies in Germany. Zhang et al. [15] studied

several freeway sites in the US and concluded that the “low speed” condition has to be sustained for at least 15 min to be recognized as a “true” breakdown instance.

Based on Elefteriadou [18], the underlying logic for identifying roadway capacity is shown by the flow chart in Fig. 2. The breakdown time duration in this case is 3 consecutive time intervals or 15 min.

As discussed above, the speed threshold and the breakdown duration time are location dependent, and consequently, it is critical that the user picks appropriate values. Intuitively, if the thresholds are too “liberal,” many “breakdowns” will be identified, and if the thresholds are too “conservative”, no breakdown will be identified. Unfortunately, there is no theoretical approach for identifying these values and engineering judgment is often used. A downside to this approach is that two researchers using the same data may choose different threshold values, which could result in different capacity estimates.

Banks et al. [19] also proposed a multi-regime traffic flow model based on the concept of breakdown. The authors proposed a two-capacity model consisting of pre-queue flow (PQF) and queue discharge flow (QDF). They observed that flow immediately downstream of the bottlenecks decreased by a small amount at the breakdown point (i.e., $PQF > QDF$), which is termed as “capacity drop” [20, 21]. Similarly, Cassidy and Hall et al. [22, 23] claimed that there was an approximately 10 % reduction in maximum flow rates after the onset of congestion. Such a flow breakdown appeared to be triggered by speed instability.

In summary, the above three capacity estimation methods (Sects. 2.2.1–2.2.3) are widely used to estimate capacity at a given location for a single day. In general, the first two methodologies (e.g., maximum flow and curve fitting) are used in planning applications, and the breakdown method is used in traffic operation applications. What is important to note is that the three methods can lead to different capacity

values. As an example, consider Fig. 3, which shows a speed–flow diagram for westbound I405 in San Diego, California, where the green dots represent the observed flow rates over 5 min time durations for 288 periods (e.g., one complete day). The largest observed flow is 2,289 veh/h/ln and is indicated as Method 1 on the graph. The purple line shows the line of best fit using the Van Aerde approach. This results in a capacity estimate of 1,774 veh/h/ln and is indicated as Van Aerde Capacity in Fig. 3. Using the definition of breakdown by Elefteriadou [18], the breakdown flow rate of 2,223 veh/h/ln can be identified. This is indicated as Breakdown Capacity in Fig. 3. This breakdown capacity point delimits the congested and uncongested parts of the observed traffic flows.

In addition, each of these approaches can be repeated across days. This would then yield a histogram of estimated capacity values and the user could use some measure of central tendency (e.g., mean or median) to identify the capacity. The following sections describe one approach for identifying capacity based on the assumption that capacity is intrinsically stochastic.

2.3 Capacity estimation method for multiple days: product limit method

Based on daily observations of traffic data collected over several months, Brilon et al. [17] argued that capacity is Weibull-distributed with a nearly constant shape parameter. Based on this observation, the authors developed the product limit method (PLM) for estimating the capacity distribution function from empirical data.

The PLM is related to (1) the breakdown flow rate, and (2) those flow rates that do not result in a breakdown occurrence. There might be certain intervals that exceed the threshold, and these are regarded as censored data (i.e., data in intervals that do not provide breakdown information). In a similar manner, traffic breakdown is regarded as a failure

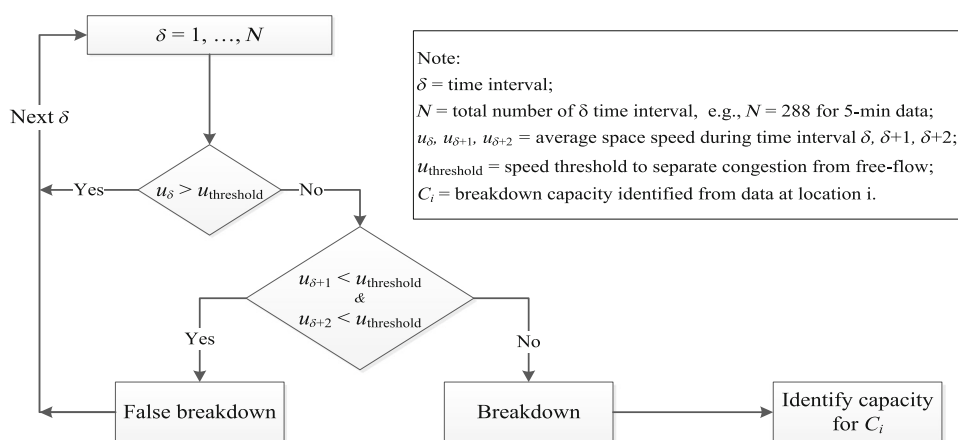


Fig. 2 Flow chart of breakdown capacity estimation

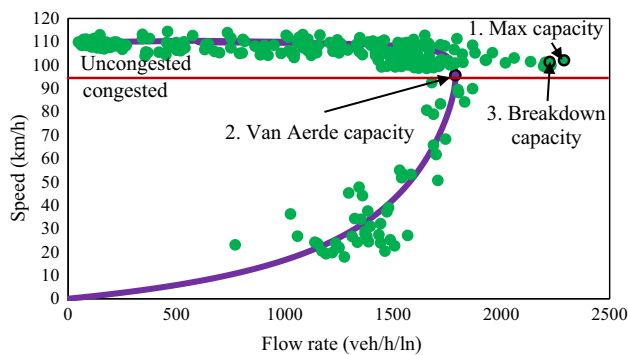


Fig. 3 Three capacity estimation methods (data in 5-min aggregation, westbound I405, San Diego)

event and it is analogous for estimating the capacity as the lifetime in a lifetime data analysis [24]. The statistics of this lifetime analysis can be then used to estimate the parameters of the distribution function, which includes the censored data. The survival function, described by the non-parametric PLM [25], is shown in Eq. (7).

$$F_c(q) = 1 - \prod_{i:q_1 \leq q} \frac{k_i - d_i}{k_i}, \quad (7)$$

where $F_c(q)$ denotes the distribution function of capacity c ; q denotes the traffic flow rate; q_i denotes the traffic flow rate in interval i ; k_i denotes the number of intervals with a traffic volume of $q \geq q_i$; and i denotes the 5 min intervals belonging to the lifetime T (i.e., a day).

The following five steps are used to define the capacity:

- (1) Identify breakdowns using flow chart in Fig. 2. Then, delete all the data in the interval that a breakdown happens. Order the 5-min interval flow and speed censored data over time.
- (2) Group the data in 15 veh/h incremental of flow levels. Count the total number of intervals y_i and the breakdown intervals d_i that fall into each group i ;
- (3) Calculate the survival probability $p_{ci} = 1 - \frac{d_i}{y_i}$ for each 15 veh/h increment of flow levels.
- (4) Calculate the overall bottleneck capacity by continuously production: $S(q_j) = \prod_{i=1}^j p_{ci}$.
- (5) Plot the survival probability curve. Choose a survival probability (e.g., 95 %), and estimate the capacity from the curve.

3 Sites and data

3.1 The study sites

The two study sites used in this paper were chosen because they contained known “bottlenecks” where breakdown

phenomena would more readily be observed [17]. The first is the westbound section of Interstate 405 (I405) in San Diego, California, as shown in Fig. 4a. It may be seen that there are four lanes and an off-ramp upstream at the detector location. The posted speed limit is 70 mi/h (110 km/h). The area is primarily urban and is often congested. Approximately 6 months of dual inductive loop data, lasting from March 3, 2013 through Sep. 27, 2013, were collected from the Caltrans Performance Measurement System (PeMS) website [26] at loop detector ID = 1217573. The 30-second flow and speed by-lane data were aggregated into 5-min intervals over the entire 24 h period (e.g., 288 5 min periods per day). A total of 112 weekdays were observed.

The second site is an eastbound section of the Inner-Ring Expressway located at the Wuning crossroad in Shanghai, China, as shown in Fig. 4b. There are two main lanes, and the posted speed limit is 80 km/h. The loop detector (ID = NHWN_40) is located approximately 102 m downstream from the Merge End, shown in Fig. 4b. The area is primarily urban, and the site experiences significant congestion. Approximately 5 months of dual inductive loop data, from June 1, 2010 through October 29, 2010, were collected. The 20-second flow and speed by-lane data were aggregated into 5-min intervals over the entire 24 h period (i.e., 288 5-min periods per day). A total of 81 weekdays were observed.

Note the two test beds were chosen because both systems have been studied extensively and have been calibrated on a regular basis [27–30]. In addition, prior to the analysis, the data were analyzed to identify detector malfunctions (e.g., missing data or abnormal data) or extraordinary events (e.g., congestion time and occupancy were order-of-magnitude larger than average). No outlier events were identified.

3.2 Preliminary data analysis

Figure 5a shows the speed and flow, aggregated to a 5 min average, as a function of time of day for the San Diego test site on April 17, 2013. It may be seen that the traffic flows fairly smoothly as evidenced by the relatively high speed experienced for the majority of the day. The exception is during the period from 14:45 to 19:05 where there is a considerable decrease in both flow rate and speed, which is an indication of congested conditions. After this period, the speed recovers to pre-congested conditions. At approximately 14:45, there is a 22 % speed drop, and this is the breakdown time identified using the logic in Fig. 2. It should be noted that this pattern is typical for the weekdays in the test dataset.

Figure 5b shows the flow and speed as a function of time at the Shanghai site on July 1, 2010. This day may be considered typical of weekdays for the data set. It may be

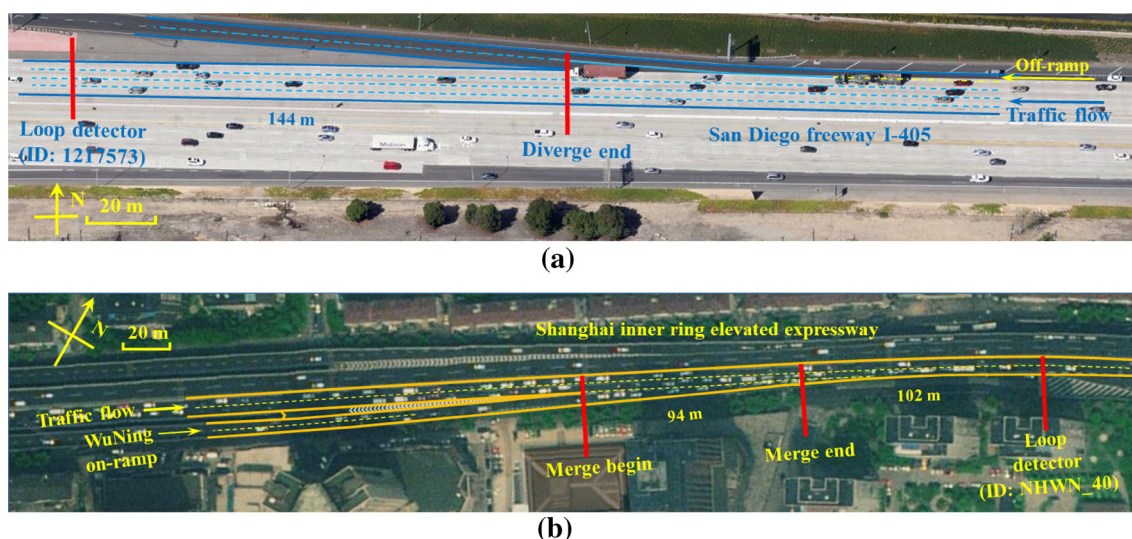


Fig. 4 The layout of the study sites. **a** Fairview Rd off-ramp, I405, San Diego site. **b** Wuning Rd on-ramp, inner-ring expressway, Shanghai site

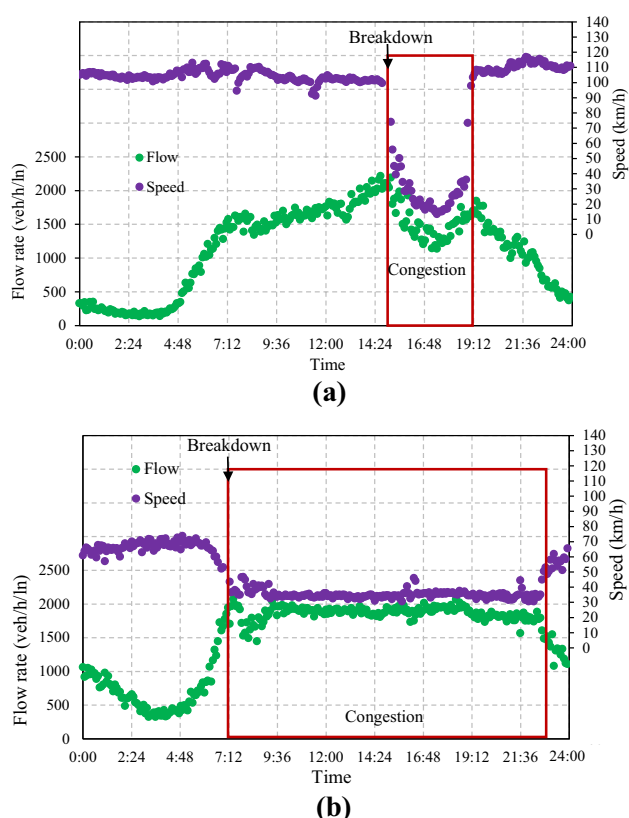


Fig. 5 Typical traffic flow and speed profiles at the two sites. **a** San Diego site, April 17th, 2013. **b** Shanghai site, July, 1st, 2010

seen that congestion lasts for approximately 16 h from 7:00 AM until approximately 11:00 PM. It can be seen in Fig. 5b that the speed gradually decreases over time starting at approximately 7:00 AM. This can be contrasted with Fig. 5a where the speed decrease is much more

abrupt. However, the breakdown identification approach shown in Fig. 2 can be used for both situations.

As discussed in Sect. 2.2.3, in order to find the breakdown point, both a threshold speed and breakdown duration need to be identified. In this study, the threshold values were based on a study of the 5-min average speed–flow data at the study sites. It was decided that 95 and 50 km/h were appropriate threshold values for the San Diego site [28] and Shanghai site [29], respectively. This paper identified a breakdown when (1) the speed decreased below the corresponding threshold, and (2) the lower speed was sustained for three consecutive 5-min intervals.

4 Capacity estimation analyses

This section first applies the maximum method, Van Aerde method, and breakdown method to estimate capacity for each day in both data sets. The resulting six capacity distributions (e.g., for both sites and all three methods) are plotted. Subsequently, the capacity over all the study days is estimated using the PLM approach for both test sites and compared with the measures of central tendency for the capacity distributions obtained by the three methods.

4.1 Capacity estimation and comparison

A preliminary analysis of the San Diego indicated the speed–flow relationships could be federated into two distinct categories. Representative days for these two categories are April 17, 2013 and June 4, 2013, and the speed–flow diagrams for these days are shown Fig. 6a and b, respectively. Note that in these figures the data is aggregated at 5 min and (1) the “Max Capacity” is the largest five minute

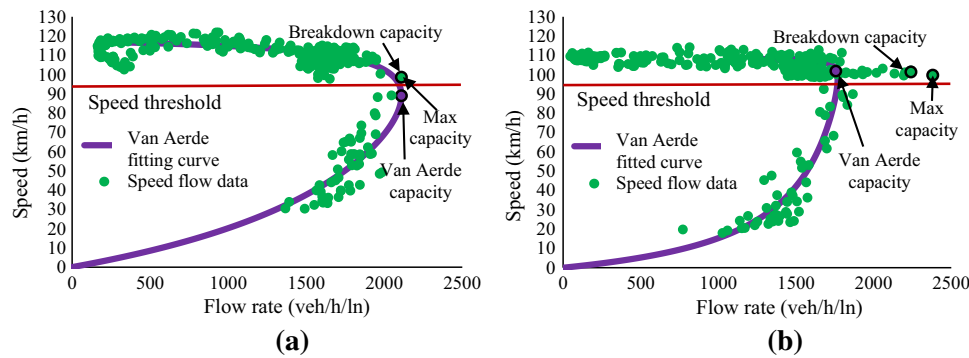


Fig. 6 Capacity estimates at San Diego site. **a** April 17th 2013. **b** June 4th 2013

flow rate observed for that particular day, (2) the “Van Aerde Capacity” is fitted by Van Aerde model from the empirical speed-flow relationship, and (3) the “Breakdown Capacity” is identified by the breakdown occurrence that is immediately prior to the start of speed drop, as shown in Fig. 2. In Fig. 6a, the capacities estimated using the maximum capacity, Van Aerde capacity, and Breakdown capacity methods are 2,109, 2,111 and 2,019 veh/h/ln, respectively. It should be noted that the differences among the three values are within 5 % of each value.

In contrast, consider the empirical data plotted in Fig. 6b which shows capacities estimated on June 4, 2013 for the San Diego site. In this case, there is a discontinuity between the uncongested and congested traffic regimes. This is also known as a capacity drop as discussed in Sect. 2.2.3. From the data in Fig. 6b, the capacities estimated using the maximum capacity, Van Aerde capacity, and Breakdown capacity methods are 2,360, 1,756 and 2,240 veh/h/ln, respectively. It should be noted that the differences among these three values are within 26 % of each estimated capacity.

Note that the speed-flow pattern shown in Fig. 6a is typical for the San Diego site in that for 82 % of the days the differences of the three capacities are within 5 % of each other. Specifically, the capacity drop phenomena, identified in Fig. 6b, were observed only on nine out of the 112 days. During these 9 days the smallest and largest differences in capacity among the three techniques were 9.7 % and 26 %, respectively.

In addition, for all 112 days at the San Diego test site the average difference between the breakdown point and the maximum capacity is 5 %. Consequently, on a typical day there is little change in average speed as the flow rate approaches its maximum value. Based on this fact, it could be argued that for this location breakdown capacity is a good approximation for maximum capacity. However, this is not always the case as will be demonstrated in the Shanghai test site analyses.

Figure 7 illustrates the relationship between speed and flow rate at the Shanghai test site on July 3, 2010. It may be

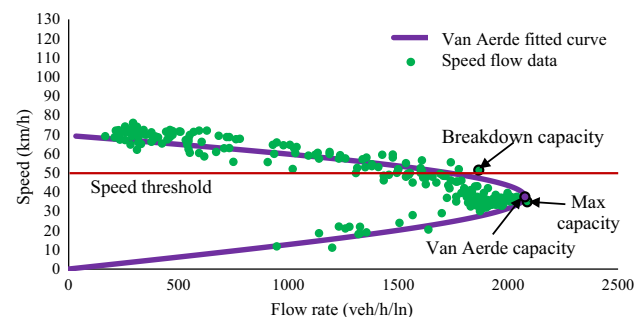


Fig. 7 Comparison of the two capacity estimates at Shanghai site, July 3rd 2010

seen that the speed gradually decreases as flow rate increases, which is in marked contrast to what was observed at the San Diego site. The pattern shown in Fig. 7 was typical for all 81 days studied. The breakdown point, identified using the logic in Fig. 2, occurs at a considerably lower flow rate than the maximum flow rate. This is referred to “early onset breakdown” by Sun et al. [29, 30]. The estimated capacities for maximum method, Van Aerde method, and breakdown method are 2,086, 2,076, and 1,866 veh/h/ln, respectively. The differences of the three values are within 11 % of each value, which is much higher than observed at the San Diego site (Fig. 6a). In addition, for all 81 days, the smallest and largest differences in capacity among the three techniques were 14 % and 34 %, respectively.

4.2 Overall capacity at each site

Histograms and boxplots of the estimated capacities at the San Diego test site for the Maximum Capacity, the Van Aerde Capacity, and the Breakdown Capacity methods are given in Fig. 8a, b, and c, respectively. It may be seen in Fig. 8a that the capacities for the maximum capacity method range from 1,760 to 2,504 veh/h/ln with a mean of 2,188 veh/h/ln. There is considerable spread in the data as evidenced by the standard deviation of 180 veh/h/ln. It can be seen in the boxplots that the distribution is skewed left

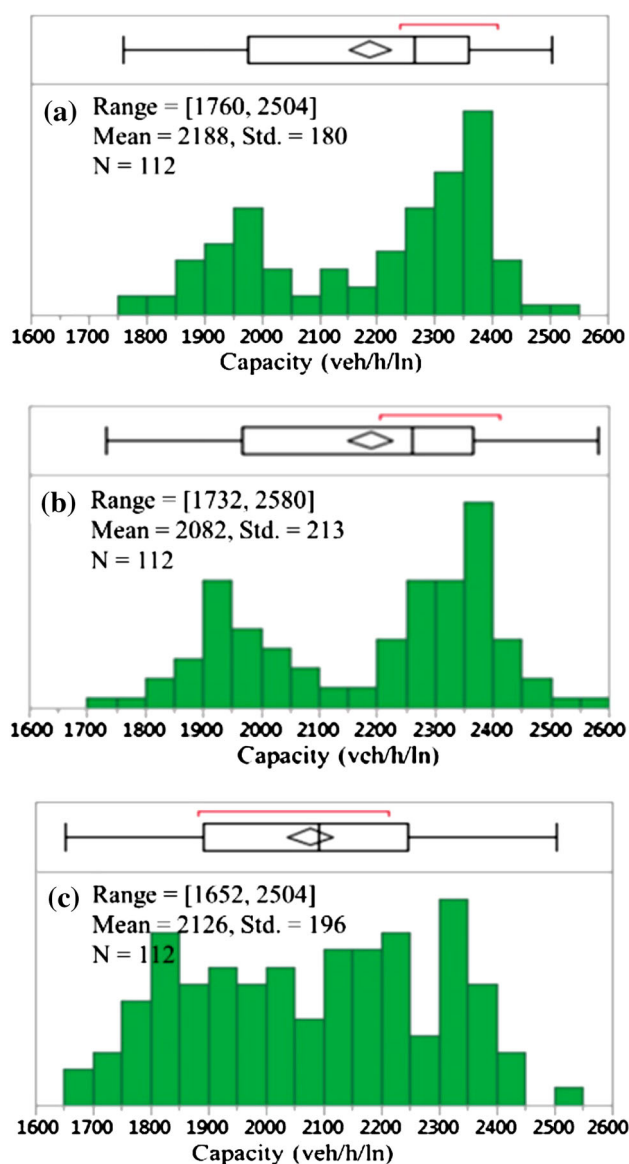


Fig. 8 Histogram and box plots of estimated capacity by maximum (a), Van Aerde (b), and breakdown (c) Methods at San Diego Test Site

with a kurtosis value of -1.1 , which indicates that more than half of the observed daily maximum capacities are higher than the mean value.

The histogram and boxplot of the capacities from the Van Aerde method are shown in Fig. 8b. It may be seen that the daily capacity has a range of 1,732–2,580 veh/h/ln, a mean of 2,082 veh/h/ln, and a standard deviation of 213 veh/h/ln. The Van Aerde capacity also has a bi-modal distribution and is also skewed left with a kurtosis value of -1.2 . It was hypothesized that this occurred because of the 9 days in which a capacity drop was observed.

The histogram and boxplot of the daily capacities from the breakdown method are shown in Fig. 8c. In contrast to the first two graphs, this distribution is unimodal and

approximately uniform, though the Kolmogorov–Smirnov test did not show significant under uniform distribution (see Table 1). The mean value is 2,126 veh/h/ln, with a median of 2,096 veh/h/ln, and both of these metrics are good indicators of the central tendency.

Histograms and boxplots of the estimated capacities at the Shanghai test site for the maximum capacity, the Van Aerde capacity, and the breakdown capacity methods are given in Fig. 9a, b, and c, respectively. It may be seen in Fig. 9a that the capacities for the maximum capacity method range from 1,923 to 2,178 veh/h/ln with a mean of 2,049 veh/h/ln. There is considerably less spread in the estimated capacities, as compared to the San Diego test site, as evidenced by the standard deviation of 56 veh/h/ln. The Van Aerde capacity has a similar distribution as seen in Fig. 9b. The daily capacity values estimated using this method range from 1,986 to 2,178 veh/h/ln with a similar mean of 2,066 veh/h/ln and a standard deviation of 47 veh/h/ln.

It may be seen in Fig. 9c that the capacities for the Breakdown Capacity method range from 1,434 to 2,128 veh/h/ln. The mean of this distribution was 1,789 veh/h/ln, which is 14 % and 13 % lower than that of the maximum method and Van Aerde method, respectively. In addition, there is considerably more spread in the estimated capacities, as compared to the maximum capacity and Van Aerde capacity method, as evidenced by the standard deviation of 146 veh/h/ln. It may be seen that the distribution is unimodal and is approximately Gaussian shaped.

The stochastic nature of the estimated capacities for each methodology is captured in Figs. 8 and 9. While a visual inspection can indicate the general form (e.g., uniform, Gaussian, etc.) for each method, it is unclear which distribution, if any, best fits the estimated values. Four test distributions, the Normal/Gaussian, Lognormal, Weibull, and Uniform, were tested using a one-sample Kolmogorov–Smirnov test. The testing was conducted at the 5 % level of significance, and the results are shown in Table 1.

The last column in Table 1 shows the “best” distribution based on the p value. It can be seen at the San Diego site none of the generic distributions was found to provide a statistically significant fit for the maximum and Van Aerde methods. This is not unexpected given the bi-modal nature of these latter distributions and the fact that the tested distributions are unimodal.

It can be seen from Figs. 8 and 9, and Table 1 that the capacities vary widely depending on test site and capacity estimation technique. For example, the capacity estimated by maximum method at San Diego is similar to a bimodal distribution. In contrast, the capacities estimated at Shanghai test site (Fig. 9) were normally distributed with a comparatively tight range.

Based on the PLM described in Sect. 2.3, the capacity value in this paper is taken to be the 5th percentile value of

Table 1 One-sample Kolmogorov–Smirnov Test (at 5 % level of significance)

	Normal	Lognormal	Weibull	Uniform	Distribution type
San Diego					
Maximum method	0.014	<0.001	0.018	0.034	None
Van Aerde method	0.003	<0.001	<0.001	<0.001	None
Breakdown method	0.016	0.010	0.076	0.029	Weibull
Shanghai					
Maximum method	0.200	0.150	<0.001	0.005	Normal/lognormal
Van Aerde method	0.007	<0.001	0.155	0.006	Weibull
Breakdown method	0.028	0.059	0.036	<0.001	Lognormal

H_0 the data is from the to-be-tested distribution. Small p values reject H_0

Italics values signifies at 5% significant level

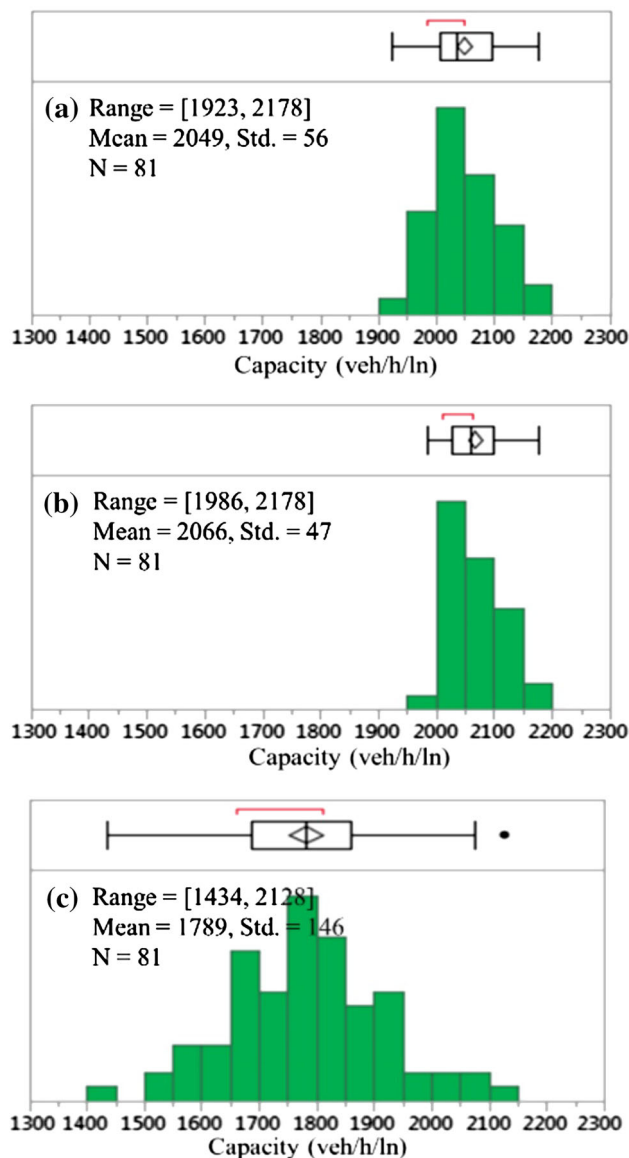


Fig. 9 Histogram and box plots of estimated capacity at Shanghai by maximum, Van Aerde and breakdown method, respectively

the cumulative distribution function breakdown [12]. This means there is 95 % probability that a bottleneck will have dispersed when the flow rate is greater than the PLM capacity as shown in Fig. 10. The PLM-based capacity is identified as 2,217 and 1,733 veh/h/ln at San Diego site and Shanghai site, respectively, as shown in Fig. 10.

The PLM capacity is compared with the capacity estimated from the maximum method, Van Aerde method, and breakdown method as shown in Table 2. At the San Diego site, the 25th and 75th percentile of capacity range estimated by all the three methods includes the PLM capacity. In contrast, at the Shanghai site, only the breakdown method is within the 25th and 75th of capacity range, while the PLM capacity does not fit in the ranges for either the maximum flow method or Van Aerde method.

The Wilcoxon signed-rank test (nonparametric test) is used to test the hypothesis that there are no significant differences between the PLM capacity and capacities estimated by the three methods. The tests were conducted with α -value equal to 0.05. In other words, a p value larger than 0.05 will mean the alternative hypothesis is rejected. As seen in Table 2, it can conclude that at the San Diego site, the capacity estimated by the PLM method is statistically the same as the capacity values obtained by the maximum method and the Van Aerde method. In contrast, it is concluded that the capacity estimated by the PLM method is statistically different than the capacity estimated by the breakdown method.

At the Shanghai site, the PLM capacity is statistically the same as the breakdown capacity as evidenced by the fact that the p value of 0.09 is greater than 0.05. The 25th–75th percentile ranges for the maximum and Van Aerde methods, however, do not include the PLM capacity. This would be expected because of the nature of the PLM approach which first identifies a breakdown and then uses only censored data to estimate capacity.

As a point of reference, HCM capacities for San Diego and Shanghai, in terms of units, are also shown in Table 2.

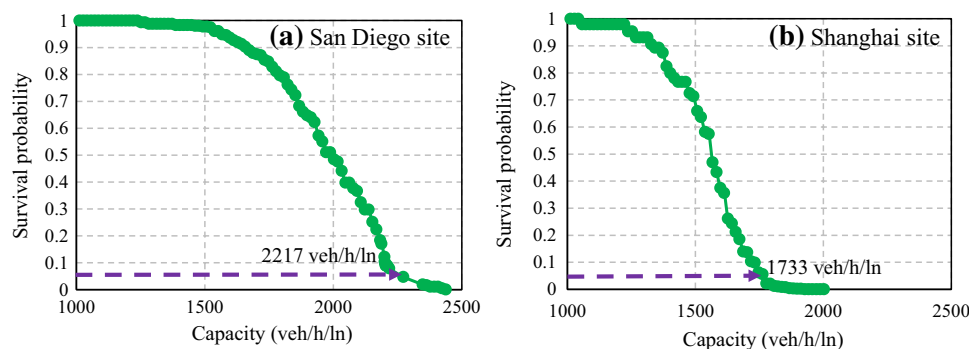


Fig. 10 Capacity estimate by PLM at each site

Table 2 Estimated capacity distribution statistics (unit: veh/h/ln)

Site	Method	Mean	Median	[25th, 75th] percentile	PLM capacity	Nonparametric test ($\alpha = 0.05$)	HCM capacity
San Diego	Maximum	2,188	2,264	[1,976, 2,360]	2,217	0.7925	2,400
	Van Aerde	2,188	2,260	[1,968, 2,364]		0.1641	
	Breakdown	2,078	2,092	[1,892, 2,231]		0.0162	
Shanghai	Maximum	2,049	2,035	[2,006, 2,097]	1,733	<0.0001	2,200
	Van Aerde	2,066	2,058	[2,028, 2,100]		<0.0001	
	Breakdown	1,789	1,782	[1,686, 1,860]		0.0910	

Unit in HCM: passenger car per hour per lane

Italic values signifies at 5% significant level

These were based on their free-flow speeds, and it should be noted that heavy vehicle factor was not accounted. Comparing the capacities obtained by the PLM methodology and the HCM methodology, it may be seen that the values are much closer at the San Diego site. In contrast, the capacity values estimated by the PLM and HCM approaches are considerably different at the Shanghai site. However, as seen in Table 2, in comparison to the HCM capacity, the maximum method and Van Aerde model provide capacity estimates that are 7 % closer.

5 Concluding remarks

A wide variety of authors have developed methodologies for estimating capacity from empirical data sets that are collected automatically as part of an ITS. These include (1) simply observing the maximum flow rate, (2) using basic curve fitting techniques based on simple assumptions related to the fundamental diagram, (3) calculating capacity estimates based on breakdown phenomena, and (4) approaches for estimating average daily capacity based on the stochastic nature of capacity and multiple days of data. Note that a comprehensive literature review is beyond the scope of this paper. Instead, the goal was to compare and contrast four of the most popular capacity estimation techniques in terms of (1) data requirements, (2) modeling

effort required, (3) estimated parameter values, (4) theoretical background, and (5) statistical differences across time and over geographically dispersed locations.

In summary, there were three major conclusions arising from these analyses. The first is that irrespective of the estimation methodology and the definition of capacity, the estimated capacity varies over time. For this case study, three methods (e.g., maximum method, Van Aerde model, and breakdown method) were used to estimate the capacity at the San Diego test site and the Shanghai test site over multiple days. It was found that at the San Diego site, the differences of the estimated capacity were within 5 % among 103 of the 112 days (e.g., 82 % of the time). For nine of the 112 days a “capacity drop” was identified, and on these days, the smallest and largest differences in estimated capacity were 9.7 % and 26 %, respectively. In contrast, at the Shanghai site the capacity estimates for the three methodologies were between 14 % and 34 % for all 81 days. There was not capacity drop observed at the Shanghai site. While the two test sites were located on major thoroughfares in two large metropolitan areas, the estimated daily capacity values were markedly different. It is hypothesized that this occurred because of differences in: (1) vehicle types/capabilities, (2) vehicle distributions, and (3) driving behavior.

It was also found that the estimated capacity at the Shanghai site was markedly different than the equivalent

HCM values. Using the HCM methodology the estimated capacity would be 2,400 and 2,200 pcphpl for free flow speeds of 110 and 80 km/h, respectively. In contrast, the estimated capacity at the Shanghai test side ranged from 1,686 to 2,100 veh/h/ln. It is hypothesized that the difference in capacity may come from the heavy vehicle factor which was ignored in this study and the fact that the HCM is based on 15 min aggregate flows, while this study utilized 5 min aggregate flows.

Regardless, it is advantageous to understand the distribution of capacities over time when selecting the “true” capacity at a site. Simply picking a day at random will, in all likelihood, lead to an erroneous result unless the traffic flow rates are homogeneous over time.

The second finding was that, not surprisingly, the estimated capacities were different for each of the techniques. In general, the maximum method focused on the maximum value of the traffic flow rates over a day, and therefore it tends to define the upper bound of the estimated capacity range. The Van Aerde method estimated the theoretical capacity value by regressing all the flow rate data over a single day. It was found that the differences of the estimated capacity values are within 5 % over all days. Capacities obtained by the breakdown method vary largely depending on the traffic situation at a particular site. Usually, this approach identifies a lower capacity than the other techniques, and this may serve as a lower bound of the estimated capacity range. The PLM estimation values typically are within the ranges identified by the first three methods. This is not surprising because the PLM estimation explicitly accounts for day by day variability.

The third finding was that each of the methodologies has advantages and disadvantages, and the best method will be function of the available data, the application, and the goals of the modeler. It was found that the maximum method is the easiest method, from a computational perspective, for estimating capacity from empirical data. However, there was very little variation in capacity estimates over time using this method, and the method cannot be used to identify major changes in flow rate over short periods of time. It was found that the Van Aerde model is not tied directly to breakdown events and thus can be used to obtain deterministic capacity over the study period. This methodology is based on the traffic flow theory and does not require the user to identify the status of traffic flow on freeway a priori. It is also easy to automate and is particularly useful for uniform traffic flow such as the San Diego site. The breakdown identification method accounted for the stochastic nature of capacity, which many authors believe that it leads to more credible results. The authors argued that it is necessary to identify the breakdown conditions, otherwise the modeler is unsure of whether a higher flow rate could be observed. Similar to the previous two models, the breakdown method is easily adapted to different

types of freeways. Although the determination of an appropriate breakdown capacity is not straightforward for many applications, it is critical to identify breakdown points, and in these situations, such as at the Shanghai site, breakdown-related approaches are very useful.

In future work, several aspects that would increase the accuracy of capacity estimation could be addressed including (1) a larger sample size with more days and multiple locations, (2) an analysis of the effect of vehicle type including percentage of heavy vehicles, (3) an analysis of the traffic flow distribution across lanes, and (4) an analysis of the traffic flow distribution in terms of the effect of weather. Lastly, it would be interesting to relate individual driver behavior characteristics under distinctive traffic situations to the capacity estimation.

Acknowledgments Special thanks are due to Dr. Jian Sun of the Tongji University in China and the PeMS in California for their assistance in providing data. The analysis and findings related to this data are strictly those of the authors and not necessarily those of the people and organizations who provided the underlying data.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Bureau of Public Roads (1950) Highway Capacity Manual. Practical Applications in Research
2. Highway Research Board (1965) Highway Capacity Manual. Washington, DC
3. Transportation Research Board (1985) Highway Capacity Manual. Washington, DC
4. Transportation Research Board (2000) Highway Capacity Manual. Washington, DC
5. Transportation Research Board (2010) Highway Capacity Manual. Washington, DC
6. Brilon W, Geistefeldt J, Zurlinden H (2007) Implementing the concept of reliability for highway capacity analysis. *Transp Res Rec* 2027(1):1–8
7. Elefteriadou L, Roger PR, William RM (1995) Probabilistic nature of breakdown at freeway merge junctions transportation research record. *Transp Res Rec* 1484:80–89
8. Persaud B, Yagar S, Brownlee R (1998) Exploration of the breakdown phenomenon in freeway traffic. *Transp Res Rec* 1634:64–69
9. Wu X, Michalopoulos P, Liu XH (2010) Stochasticity of freeway operational capacity and chance-constrained ramp metering. *Transp Res Part C* 18:741–756
10. Minderhoud MM, Botma H, Bovy PH (1997) Assessment of roadway capacity estimation methods. *Transp Res Rec* 1572(1):59–67
11. Geistefeldt J, Brilon W (2009) A comparative assessment of stochastic capacity estimation methods. *Transportation and Traffic Theory: Golden Jubilee*. Springer US, pp 583–602

12. Chang M, Kim Y (2000) Development of capacity estimation method from statistical distribution of observed traffic flow. Transportation Research Board, Transportation Research Circular, pp 299–309
13. Zunhwan HW, Jumsan K, Sungmo R (2005) Development of a new highway capacity estimation method. In: Proceedings of the Eastern Asia society for transportation studies, Vol 5
14. Greenshields BD, Channing W, Miller H (1935) A study of traffic capacity. In: Proceedings of the highway research board, Washington DC
15. Van Aerde M (1995) Single regime speed–flow–density relationship for congested and uncongested highways. Paper presented at the 74th TRB Annual Conference, Washington DC
16. Zhang L, Levinson D (2004) Some properties of flows at freeway bottlenecks. *Transp Res Rec* 1883:122–131
17. Brilon W, Geistefeldt J, Regler M (2005) Reliability of freeway traffic flow: a stochastic concept of capacity. In: Proceedings of 16th international symposium of transportation and traffic theory, University of Maryland, College Park, pp 125–144
18. Elefteriadou L, Lertworawanich P (2003) Defining, measuring and estimating freeway capacity. Paper presented at the 82nd TRB Annual Conference, Washington DC
19. Lorenz MR, Elefteriadou L (2001) Defining freeway capacity as function of breakdown probability. *Transp Res Rec* 1776(1): 43–51
20. Banks JH (1991) The two-capacity phenomenon: some theoretical issues. *Transp Res Rec* 1320:234–241
21. Banks JH (2006) New approach to bottleneck capacity analysis: Final Report UCB-ITS-PRR-2006-13, California PATH. University of California, Berkeley 2006
22. Cassidy ML, Bertini RL (1999) Some traffic features at freeway bottlenecks. *Transp Res Part B* 33(1):25–42
23. Hall FL, Agyemang-Duah K (1991) Freeway capacity drop and the definition of capacity. *Transp Res Rec* 1320:91–98
24. Brilon W, Geistefeldt J, Zurlinden H (2007) Implementing the concept of reliability for highway capacity analysis. *Transp Res Rec* 2027(1):1–8
25. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 53:457–481
26. PeMS Data Resource Website: <http://pems.dot.ca.gov/>. Accessed 10 May 2014
27. Chen C et al (2001) Freeway performance measurement system: mining loop detector data. *Trans Res Rec* 1748(1):96–102
28. Jia Z, et al. (2001) The PEMS algorithms for accurate, real-time estimates of g-factors and speeds from single-loop detectors. In: Proceedings of the Intelligent Transportation Systems. IEEE, 2001
29. Sun J, Zhang J, Zhang HM (2013) Investigation of early-onset breakdown phenomenon at urban expressway bottlenecks in Shanghai. Paper presented at the 92nd TRB Annual Conference, Washington DC
30. Sun J, Zhao L, Zhang HM (2014) The mechanism of early-onset breakdown at Shanghai's expressway on-ramp bottlenecks. *Transportation Research Record* 3516